Biophysical Journal Volume 105 December 2013 2598–2605

The Role of Semidisorder in Temperature Adaptation of Bacterial FlgM Proteins

Jihua Wang,[†‡] Yuedong Yang,[§¶||††] Zanxia Cao,[†‡] Zhixiu Li,[§¶] Huiying Zhao,[§¶] and Yaoqi Zhou[†§¶||††*]
[†]Shandong Provincial Key Laboratory of Functional Macromolecular Biophysics and [‡]School of Physics and Electronic Information, Dezhou University, Dezhou, Shandong Province China; [§]Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana; and [¶]School of Informatics, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana; [||]Institute for Glycomics and [††]School of Information and Communication Technology, Griffith University, Southport, Australia

ABSTRACT Probabilities of disorder for FlgM proteins of 39 species whose optimal growth temperature ranges from 273 K (0°C) to 368 K (95°C) were predicted by a newly developed method called Sequence-based Prediction with Integrated NEural networks for Disorder (SPINE-D). We showed that the temperature-dependent behavior of FlgM proteins could be separated into two subgroups according to their sequence lengths. Only shorter sequences evolved to adapt to high temperatures (>318 K or 45°C). Their ability to adapt to high temperatures was achieved through a transition from a fully disordered state with little secondary structure to a semidisordered state with high predicted helical probability at the N-terminal region. The predicted results are consistent with available experimental data. An analysis of all orthologous protein families in 39 species suggests that such a transition from a fully disordered state to semidisordered and/or ordered states is one of the strategies employed by nature for adaptation to high temperatures.

INTRODUCTION

Intrinsically disordered proteins (IDPs) form a class of proteins that lack a stable tertiary and/or secondary structure under physiological conditions (1,2). IDPs play diverse functional roles in many biological processes. Their functions include regulation of transcription and translation, cellular signal transduction, protein phosphorylation, storage of small molecules, protein-nucleic-acid recognition, regulation through degradation, and regulation of self-assembly of large multiprotein complexes such as bacterial flagellum (3,4). Over the past decade, there has been a growing recognition that ~10–35% of prokaryotic and 15–45% of eukaryotic proteins likely contain at least one intrinsically disordered region >30 residues in length. Moreover, new functional roles of IDPs are being uncovered (1,5–7). Advances in experimental characterization of IDPs have been made with NMR spectroscopy, small-angle x-ray scattering, and single-molecule techniques (7). Meanwhile, more than 50 predictors of intrinsic disorder have been developed (8,9). Molecular-dynamics simulations in combination with experimental data have also been applied to depict the heterogeneous nature of the conformations of IDPs (10–12).

One area of particular interest is the role of intrinsic disorder in temperature adaptation of proteins. Studies of psychrophilic, mesophilic, and thermophilic enzymes (13–15) suggest that cold adaptation leads to more disorder to maintain dynamics and function, whereas thermophilic enzymes are more structured (ordered). Genome-scale analyses of structural disorder in thermal adaptation of prokaryotes indicate that a significant reduction in structural disorder is accompanied by a reduction in genome and proteome size in adaptation to habitats at very high temperatures (16,17).

The above analyses assumed that all IDPs are the same. However, in fact, different disordered proteins behave differently; their conformations can be characterized as random coils, semicompact, or compact globules (called premolten or molten globules) with varying content of secondary structures (18,19). Some disordered proteins can fold when interacting with their respective partners, whereas others do not (19). Some disordered proteins are aggregation prone, and others are aggregation resistant (20). More importantly, theoretical studies of protein-like polymers indicate the existence of coil, disordered globules, and ordered globules (21) in addition to native-like surface-molten solids (22). Thus, it is of interest to conduct an in-depth analysis of the roles played by different types of disorder in temperature adaptation of IDPs.

Recently, we developed a method called Sequence-based Prediction with Integrated NEural networks for Disorder (SPINE-D) (9). The method was ranked as one of the best techniques for predicting protein disorder in Critical Assessment of Structure Prediction (CASP 9) (23). More interestingly, the method is capable of defining a physically meaningful state of semidisorder with a disorder probability between 0.4 and 0.7 (24). A semidisordered state is partially collapsed, with some secondary structures, and more likely to be involved in induced folding and protein aggregation (24).

In this work, we focus on temperature adaptation of bacterial FlgM proteins. FlgM proteins are negative regulators for

Submitted September 2, 2013, and accepted for publication October 18, 2013.

*Correspondence: yqzhou@iupui.edu or yaoqi.zhou@griffith.edu.au

Editor: David Eliezer.

0006-3495/13/12/2598/8 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.10.026

synthesis of flagellin proteins that form a long helical filament that is important for bacterial movement in a liquid environment (25,26). NMR studies of free FlgM from *Salmonella typhimurium* at 25°C indicated that it is mostly unstructured in a dilute solution, but its C-terminal half forms a transient helix (27). The disordered FlgM protein can gain its structure under biologically relevant conditions, such as in solutions containing high concentrations of glucose (28), suggesting the role of semidisorder. On the other hand, the corresponding thermophilic protein from *Aquifex aeolicus* has more secondary structure content (29) and the structure of its functional complex (PDB ID 1RP3) was determined by x-ray crystallography at 2.3 Å resolution (30). Further experimental studies of more FlgM proteins adapted at different temperatures revealed varying degrees of compactness and secondary structural content (31).

Here, we performed a detailed analysis of 39 homologous FlgM proteins from different bacteria selected from UniProt (32) and PGTdb (a database that provides growth temperatures of prokaryotes) (33). The disorder probabilities of these FlgM proteins were predicted by SPINE-D (9). We show that semidisordered regions of thermophilic FlgM have high helical probabilities. In contrast, a fully disordered region for mesophilic FlgM does not have any secondary structure. Our results highlight the important role of semidisorder in high-temperature adaptation of FlgM proteins.

MATERIALS AND METHODS

Data set for FlgM proteins

To construct a data set of FlgM proteins from species with different optimal growth temperatures, we first searched all FlgM proteins in UniProt (32) and obtained 914 FlgM proteins from different organisms. We kept only those proteins in organisms with records of optimal growth temperature in the PGTdb (33). This led to 39 FlgM proteins as shown in Table 1.

Multiple sequence alignment

We employed T-coffee (34) to carry out multiple-sequence alignment of 39 FlgM proteins and ensure that equivalent sequence positions were obtained to facilitate comparison of disordered regions in different proteins.

Disorder and secondary structure predictors: SPINE-D and SPINE-X

We employed SPINE-D (9) because it is capable of separating a semidisordered state from a fully disordered state (24). The SPINE-D server (http://sparks-lab.org) was employed to predict the disorder probability of 39 homologous FlgM proteins. We also employed the recently developed method SPINE-X (35,36) for secondary structure prediction because it is one of the most accurate predictors according to benchmark studies (82% 10-fold cross-validated accuracy for three-state prediction).

Evolution tree server: RAxML

To understand the evolution relation of 39 homologous FlgM proteins, it is necessary to construct the evolution tree of FlgM proteins. In this work, we employed RAxML to map the evolution tree (37) (http://phylobench.vital-it.ch/raxml-bb/index.php).

Analysis of other orthologous proteins

To go beyond the results of the family of FlgM proteins, we obtained all proteins from 39 genomes and removed redundant proteins with >30% sequence identity within each genome with Blastclust (38). Then, we randomly selected mesophilic species (*Desulfovibrio vulgaris*, optimal growth temperature of 310 K (37°C)) and searched for orthologous proteins of *D. vulgaris* in the other 38 species with PSI-BLAST (39) with E-value < 0.001. Only proteins with orthologs in all 38 species (a total of 572×39 proteins) were studied in this work. SPINE-D was utilized to predict order, semidisorder, and full disorder for the 572×39 proteins. We further defined disordered regions according to *D. vulgaris* proteins with a minimum of 10 continuous residues with a predicted disorder probability ≥ 0.4. The corresponding amino acid residues for the disordered regions in *D. vulgaris* proteins were obtained by multiple-sequence alignment of orthologous families. The average disorder probabilities for all disordered regions in a single protein were obtained for all 527 proteins in the 39 species. Regions with fewer than five aligned residues were excluded from the calculations. We evaluated Pearson correlation coefficients between average disorder probabilities and optimal growth temperatures if more than 34 species had the region aligned with five or more residues to the region in *D. vulgaris* proteins.

RESULTS

Predicted disorder probabilities

Fig. 1 shows disorder probabilities obtained from SPINE-D for 39 FlgM proteins based on their aligned positions so that their ordered or disordered regions could be displayed within the same figure. For the majority of proteins, only the ~25-residue-long region around residue 100 in aligned positions has a strong tendency to form structures (disorder probability < 0.4). However, the tendency is lower for several proteins. This region also corresponds to the most conserved region, highlighting the importance of this structure-forming ordered region for the function of FlgM in all species.

In addition to ordered regions, many proteins have semidisordered (0.4–0.7) and fully disordered (>0.7 in disorder probability) regions. The most visible effect of the high-temperature adaptation, however, is the change in the N-terminal region from full disorder for psychrophilic to semidisorder for hyperthermophilic proteins. Interestingly, there is a subgroup of proteins whose N-terminal regions have a disordered probability consistently close to one. A further analysis indicates that this behavior is dependent on the size of the given protein (the number of amino acid residues).

Fig. 2 shows the disorder probabilities of the proteins separated into two subgroups based on each protein's size (subgroup A with size > 97 containing 21 proteins and subgroup B with size ≤ 97 containing 18 proteins). Only the larger proteins (subgroup A) have a consistent, fully disordered region (probability close to one) ~55 residues long, regardless of the corresponding optimal growth temperatures of the species to which they belong. In addition, they

TABLE 1 FlgM proteins listed in descending order according to optimal growth temperatures of the corresponding species

	Entry	Entry Name	FlgM Proteins	Organism	L (aa)	T (K) (°C)
1	O66683	O66683_AQUAE	Anti-sigma factor	*Aquifex aeolicus*	88	368 (95)
2	E8UWB5	E8UWB5_THEBF	Anti-sigma-28 factor FlgM family protein	*Thermoanaerobacter brockii subsp*	91	348 (75)
3	F8CXT6	F8CXT6_BACTR	Anti-sigma-28 factor FlgM family protein	*Geobacillus thermoglucosidasius*	87	341 (68)
4	E6UTF4	E6UTF4_CLOTL	Anti-sigma-28 factor FlgM family protein	*Clostridium thermocellum*	97	338 (65)
5	Q2RKH7	Q2RKH7_MOOTA	Putative anti-sigma-28 factor	*Moorella thermoacetica*	96	333 (60)
6	I3E1Z1	I3E1Z1_BACMT	Anti-sigma-28 factor FlgM family protein	*Bacillus methanolicus* PB1	88	333 (60)
7	Q65EB1	Q65EB1_BACLD	Anti-sigma factor repressor of sigma-D-dependent transcription	*Bacillus licheniformis*	88	328 (55)
8	A0PZZ1	A0PZZ1_CLONN	Regulator of flagellin synthesis	*Clostridium novyi*	86	318 (45)
9	B8KBZ7	B8KBZ7_VIBPH	Negative regulator of flagellin synthesis	*Vibrio parahaemolyticus*	104	316 (43)
10	G2UFP3	G2UFP3_PSEAI	Anti-sigma 28 factor	*Pseudomonas aeruginosa*	107	314 (41)
11	A4VJC4	A4VJC4_PSEU5	Negative regulator of flagellin synthesis	*Pseudomonas stutzeri*	109	314 (41)
12	I0GTU1	I0GTU1_SELRL	Putative anti-sigma-28 factor	*Selenomonas ruminantium*	92	313 (40)
13	A5I5H6	A5I5H6_CLOBH	Negative regulator of flagellin synthesis	*Clostridium botulinum*	93	313 (40)
14	A7Z9B2	A7Z9B2_BACA2	FlgM	*Bacillus amyloliquefaciens*	88	313 (40)
15	A3GVB3	A3GVB3_VIBCL	Negative regulator of flagellin synthesis	*Vibrio cholerae*	107	313 (40)
16	D1TT58	D1TT58_YERPE	Flagellar biosynthesis anti-sigma factor	*Yersinia pestis*	88	310 (37)
17	A5MZ39	A5MZ39_CLOK5	FlgM	*Clostridium kluyveri*	90	310 (37)
18	I2ZAT9	I2ZAT9_ECOLX	Flagellar biosynthesis anti-sigma factor	*Escherichia coli*	92	310 (37)
19	F5ZQ02	F5ZQ02_SALTU	Anti-sigma-28 factor FlgM	*Salmonella typhimurium*	97	310 (37)
20	B3WZ33	B3WZ33_SHIDY	Negative regulator of flagellin synthesis FlgM	*Shigella dysenteriae*	97	310 (37)
21	Q72EP6	Q72EP6_DESVH	Negative regulator of flagellin synthesis FlgM	*Desulfovibrio vulgaris*	104	310 (37)
22	E0T4A7	E0T4A7_EDWTF	Negative regulator of flagellin synthesis FlgM	*Edwardsiella tarda*	98	309 (36)
23	G8MAG	G8MAG0_9BURK	Anti-sigma-28 factor	*Burkholderia sp.*	108	308 (35)
24	Q3J1X7	Q3J1X7_RHOS4	Putative FlgM	*Rhodobacter sphaeroides*	104	307 (34)
25	H8FBF6	H8FBF6_XANCI	Anti-sigma-28 factor, FlgM family protein	*Xanthomonas citri pv*	103	303 (30)
26	Q5E3M0	Q5E3M0_VIBF1	Flagellar anti-sigma-28 factor FlgM	*Vibrio fischeri*	103	303 (30)
27	A9ABN9	A9ABN9_BURM1	Anti-sigma-28 factor, FlgM	*Burkholderia multivorans*	112	303 (30)
28	Q8GLQ2	Q8GLQ2_AERHY	FlgM	*Aeromonas hydrophila*	106	301 (28)
29	C5AEH5	C5AEH5_BURGB	Negative regulator of flagellin synthesis FlgM	*Burkholderia glumae*	111	300 (27)
30	C8BKB5	C8BKB5_YERRU	Anti-sigma28 factor FlgM	*Yersinia ruckeri*	98	299 (26)
31	F2ERE4	F2ERE4_PANAA	Negative regulator of flagellin synthesis FlgM	*Pantoea ananatis*	100	299 (26)
32	F3BHV0	F3BHV0_PSEHA	Negative regulator of flagellin synthesis FlgM	*Pseudoalteromonas haloplanktis*	106	299 (26)
33	G7EGR0	G7EGR0_9GAMM	Negative regulator of flagellin synthesis FlgM	*Pseudoalteromonas sp.*	105	295 (22)
34	Q485Q1	Q485Q1_COLP3	Negative regulator of flagellin synthesis FlgM	*Colwellia psychrerythraea*	102	295 (22)
35	A9DA83	A9DA83_9GAMM	Negative regulator of flagellin synthesis FlgM	*Shewanella benthica*	107	288 (15)
36	A8FHX9	A8FHX9_BACP	Transcriptional repressor FlgM	*Bacillus pumilus*	87	283 (10)
37	Q6LTR7	Q6LTR7_PHOPR	Putative negative regulator of flagellin synthesis	*Photobacterium profundum*	104	283 (10)
38	G0AU13	G0AU13_9GAMM	Flagellar biosynthesis anti-sigma factor protein FlgM	*Shewanella baltica*	93	277 (4)
39	B1XWK0	B1XWK0_LEPCP	Anti-sigma-28 factor, FlgM	*Leptothrix cholodnii*	106	273 (0)

have a semidisordered region between the 55th and 95th aligned residue positions. This semidisordered region has a large fluctuation in disorder probability, whereas changes in the disorder probabilities of fully disordered and ordered regions are smaller. Optimal temperatures for these species range from 273 K (0°C) to 316 K (43°C); that is, this subgroup of larger proteins excludes all thermophiles (318 K (45°C) or higher) and contains two psychrophiles.

For the smaller proteins (subgroup B), there is a reduction of disorder probability at N-terminal regions from near full disorder for six mesophiles and two psychrophiles to semidisorder for thermophiles as the temperature increases. The correlation coefficient between the average disorder probability in the N-terminal region and the optimal growth temperature for the 18 smaller proteins is −0.57, as shown in Fig. 3. By comparison, the corresponding correlation for subgroup A is weak (with a correlation coefficient of −0.23). This indicates that, at least for this family of proteins, semidisordered regions made high-temperature adaptation possible.

The separation of two subgroups is consistent with the evolution tree for 39 FlgM proteins. We built the evolution tree by using RAxML (37). As shown in Fig. 4, the top of the tree is dominated by subgroup B and the bottom is dominated by subgroup A. There are a few exceptions, likely due to error in construction of the tree or to the arbitrary cutoff of 97 residues used to separate the two subgroups.

Predicted secondary structure contents

Fig. 5 compares predicted secondary structures by SPINE X (35,36) for ordered and N-terminal regions of FlgM proteins in subgroups A and B. The ordered regions of all proteins (subgroups A and B) have high probabilities to form helices

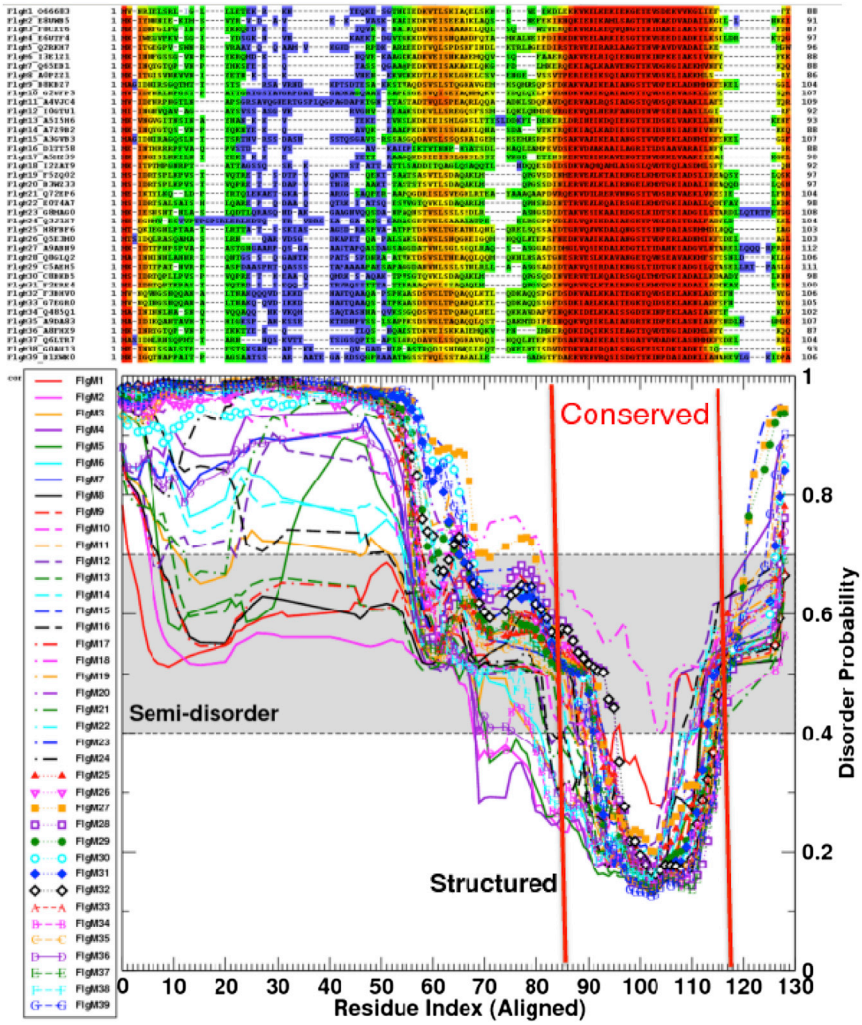

FIGURE 1 Predicted disorder probability as a function of residue indices according to the multiple-sequence alignment shown in the top panel (*red* indicates highly conserved regions). The defined semidisordered area (0.4–0.7 in disorder probability) is in gray. To see this figure in color, go online.

regardless of the magnitudes of their optimal growth temperatures, consistent with their conservation profiles. On the other hand, predicted helical probabilities in the N-terminal region for subgroup B (smaller proteins) become higher as the temperature increases, with a positive correlation coefficient of 0.55. This suggests that the change from full disorder to semidisorder is associated with an increase in helical contents. For subgroup A (larger proteins), there is no significant correlation. Only one protein has a >40% average helical content (FlgM21), corresponding to the only protein with a short semidisorder region (*green dash-dotted line*, Fig. 2 *A*). SPINE X did not predict a high probability of sheet formations in either ordered or disordered regions.

As an example, we show the predicted disorder probability and helical probability for *A. aeolicus* FlgM (FlgM1) and *S. typhimurium* FlgM (FlgM19) in Fig. 6. Thermophilic FlgM 1, dominated by semidisordered residues, has four segments with a high predicted helical probability of ~90%, in close agreement with the four helical structures when it binds with sigma-28 (PDB ID 1SC5 (30)). FlgM 19, on the other hand, is predicted to have three helices with 40 residues fully disordered, without significant secondary structure content at the N-terminal. Its structure has not yet been solved in isolation or in complex with its binding partner.

Other orthologous families

One interesting question is, do other proteins also take advantage of semidisorder for temperature adaptation? To answer this question, we calculated the correlation coefficient between average disorder probabilities for aligned regions of different species and their optimal growth temperatures. Fig. 7 *A* displays the distribution of correlation coefficients between the average disorder probability and optimal growth temperature for 327 protein families. The figure shows that the majority have weak negative correlation coefficients, but some have very strong negative correlation coefficients, similar to what we found for FlgM.

Fig. 7 *B* displays one example of a strong negative correlation coefficient between temperature and disorder

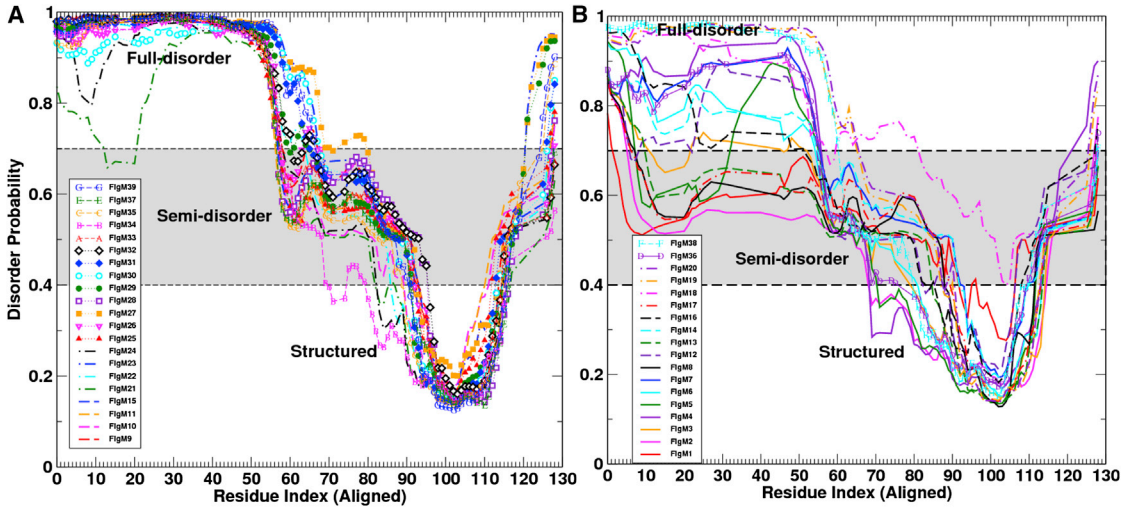

FIGURE 2 (*A* and *B*) As in Figure 1, but for two subgroups of proteins with different lengths (>97 for *A* and ≤97 for *B*). These two subgroups of proteins have distinctly different temperature-dependent disordered probabilities. The defined semidisordered area (0.4–0.7 in disorder probability) is in gray. To see this figure in color, go online.

probability, for a protein called ATP-dependent RNA helicase. It contains three disordered regions with a total of 187 residues. In this example, the fully disordered region transits into a semidisordered or fully ordered region as the optimal growth temperature increases.

DISCUSSION

In this work, we studied the change of disordered states in FlgM proteins for their adaptation to high growth temperatures. We found that formation of the semidisordered state at the N-terminal region of smaller proteins (≤97 residues) is key for adaptation to a high-temperature environment. This semidisordered state evolved from a fully disordered state (Fig. 3) that lacks secondary structure to a state that has high helical propensity (Figs. 5 and 6). A strong negative correlation between average disorder probability and optimal growth temperature was observed

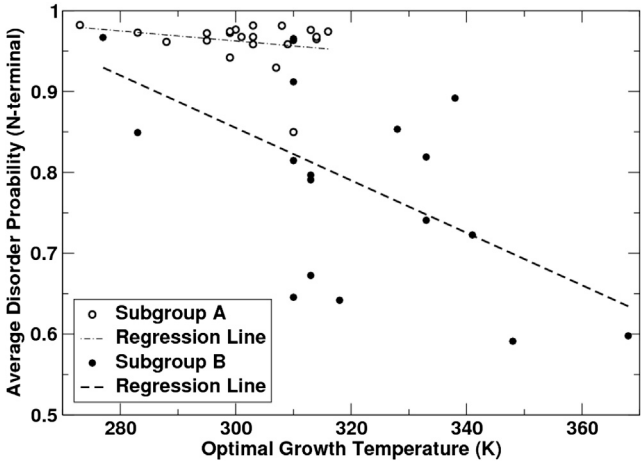

FIGURE 3 Average disorder probability for the N-terminal region (up to aligned position 56) as a function of the optimal growth temperature for proteins from subgroup A (*open circles*) and subgroup B (*solid circles*).

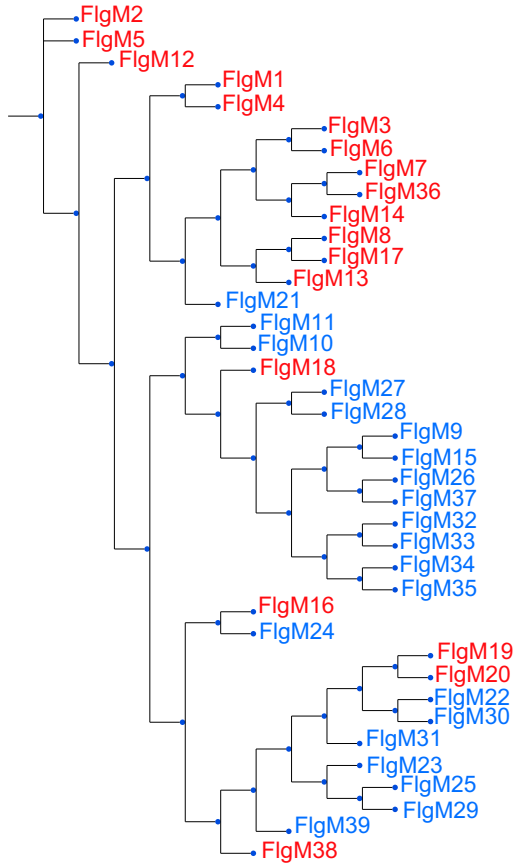

FIGURE 4 Evolution tree (*blue*, subgroup A; *red*, subgroup B). To see this figure in color, go online.

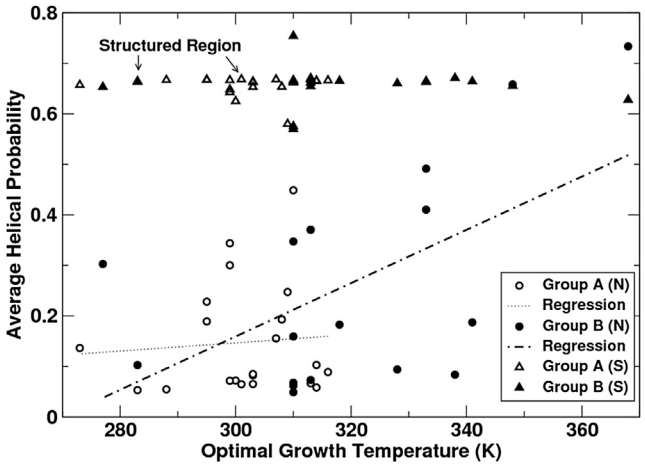

FIGURE 5 Average predicted helical probability as a function of optimal growth temperature for the N-terminal and structured regions of FlgM of subgroups A and B, respectively. Correlation coefficients are 0.08 for subgroup A (N-terminal) and 0.55 for subgroup B (N-terminal).

for smaller proteins that are capable of growth at a high temperature. Analysis of other orthologous proteins in 39 species revealed that the use of semidisorder is one of the techniques employed by nature for high-temperature adaptation.

Here, a semidisordered region in a protein refers to a region with predicted disorder probability at ~0.5 by SPINE-D (24). In our previous work (24), we demonstrated that semidisordered regions in proteins constitute a physically meaningful state because they are found to associate with induced folding and protein aggregation in IDPs and

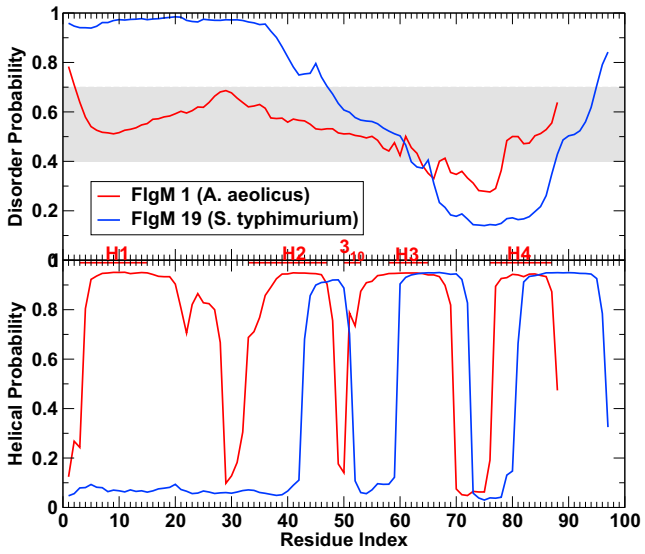

FIGURE 6 Predicted disorder probability (*top*) and helical probability (*bottom*) for thermophilic FlgM1 (*A. aeolicus* in *red*) and mesophilic FlgM19 (*S. typhimurium* in *blue*). Experimentally determined helical regions for FlgM1 in complex with sigma-28 (PDB ID 1SC5, chain B) are shown on the *x* axis with red labels. To see this figure in color, go online.

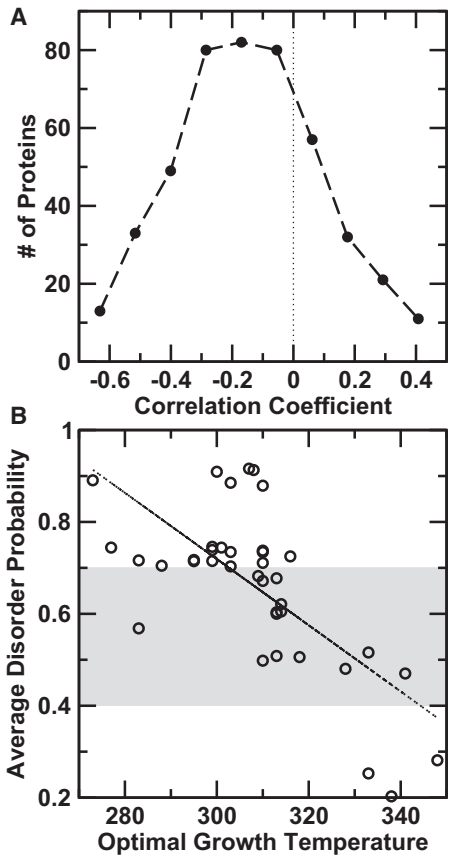

FIGURE 7 (*A*) Distribution of correlation coefficients between the disordered probabilities and temperature. The dotted line highlights the overall negative correlation observed for the majority of protein families. (*B*) Average disorder probability of ATP-dependent RNA helicases in each species along with its optimal growth temperature. The dotted line is the regression line with a correlation coefficient of −0.69. The defined semidisordered area (disorder probability of 0.4–0.7) is in gray.

locally unfolded regions in ordered proteins. To our knowledge, although the concept of semidisorder is new, the concept of the existence of multiple intrinsically disordered states is not new. Other proposed concepts are the protein trinity (order, collapsed disorder, and extended disorder) (42) and the protein quartet (folded structure, molten globule, premolten globule, and coil) (2). These concepts correspond to surface-molten solid, ordered globule, disordered globule, and coil states found in molecular-dynamics simulation studies of model proteins (21). However, how the three states (order, semidisorder, and full disorder) predicted by SPINE-D are related to the other proposed states remains to be determined in further studies, because all of these concepts are built on different computational or experimental approaches.

The predicted disordered properties of FlgM proteins with different optimal growth temperatures are consistent with known experimental data. The FlgM protein of thermophile *A. aeolicus* is the only protein whose structure was determined experimentally in the presence of its binding

partner, the flagellar sigma (30). The predicted secondary structure (four helices) is in good agreement with the observed secondary structure (Fig. 6). An NMR experimental study of the FlgM protein of mesophile *S. typhimurium* indicated that 38 amino acids residues in the N-terminal region are very flexible (disordered) even at the bound state, whereas the 24 residues within the C-terminal region possess some structure (43). This result is in close agreement with the predicted disorder probability for the protein (fully disordered in the N-terminal region and semidisordered and ordered in the C-terminal region; Fig. 6). As predicted by SPINE-X, the unbound *A. aeolicus* FlgM contains significantly more helical character in dilute solution conditions than *S. typhimurium* FlgM (28–30). Our work is also consistent with previous experimental studies of other proteins that indicated that higher disorder is involved in cold-adapted biological systems (44), and reduced disorder is important for high-temperature adaptation (13–15). The consistency between experimental data and SPINE-D predictions confirms the usefulness of disorder probabilities predicted by SPINE-D for interpreting temperature adaptation.

Our results are also consistent with previous computational studies. Several disorder predictors indicate that proteins that adapted to higher temperatures tend to have more ordered residues based on the frequency of their disorder or disorder content (16,45). We also attempted to correlate temperature with the disorder content of FlgM proteins. For example, the correlation coefficient for smaller proteins in subgroup B is only -0.20 between disorder content in the whole protein and the optimal temperature. There is no correlation if the disorder content is evaluated for the N-terminal region only, because the disorder content will be 100% in the region for all 18 proteins because their disorder probabilities in the N-terminal region are >0.5. By comparison, the correlation coefficient is -0.57 between the average disorder probability in the N-terminal region and the optimal temperature. Thus, at least for FlgM proteins, we find that the disorder probability is a more suitable parameter for correlation with optimal growth temperature. Nevertheless, the disorder content from genome-scale studies yielded results similar to those we obtained based on disorder probabilities. That is, higher growth temperatures are associated with more ordered proteins. For FlgM proteins, it is the disorder probability, rather than the disorder content, that is optimized for adaptation of FlgM proteins at different temperatures.

The disorder probability likely is not the only parameter that has been evolutionarily optimized. We have shown that different proteins evolved differently for adaptation to higher temperatures (Fig. 7). Even for FlgM, there are two subgroups. The first one, containing proteins up to 97 residues in length, can adapt to extreme high temperatures of ~360 K (87°C) by exploiting a semidisordered state. The other subgroup (with proteins >97 residues long) contains only psychrophilic and mesophilic proteins with a fully disordered N-terminal region. The separation of two subgroups based on 97 residues is somewhat arbitrary. In fact, there is no obvious separation from multiple sequence alignment (Fig. 1). Such a separation, however, does exist on the level of disorder profiles predicted by SPINE-D, as demonstrated in Fig. 2. This grouping is also supported by the evolution tree, whose top (Fig. 4) consists entirely of smaller proteins. There are a few smaller proteins scattered at the bottom, indicating that stochastic mutations can also lead to smaller proteins. This size-dependent temperature adaptation is also consistent with previous studies that showed a negative correlation between growth temperatures and genome sizes, and between growth temperatures and proteome sizes (16,17). Fig. 2 indicates that the size of a protein and its disorder probability are optimized for high growth temperatures of FlgM proteins.

Furthermore, there is an anticorrelation between the predicted disorder probability and secondary structure propensity, with a correlation coefficient of -0.54 for 39 proteins. That is, the higher the predicted disorder probabilities, the lower are the probabilities of secondary structures (helices or sheet). This is consistent with our previous finding that a semidisordered state is semistructured. The association between predicted disorder and secondary-structure probability makes it difficult to conclude whether the disorder probability or secondary-structure probability, or both, were optimized for temperature adaptation. For subgroup B of FlgM, the correlation coefficients between the average disorder probability and growth temperature, and between the average helical probability and growth temperature are -0.57 and 0.55, respectively. The correlation coefficient between the growth temperature and combined disorder and helical probability via regression is 0.61. It is possible that conformational flexibility and secondary-structure stability were both optimized for high growth temperatures.

In summary, this work suggests that proteins do not have to be fully ordered to adapt to high temperatures. Enhancing the structural stability of functional complexes through induced folding while maintaining the dynamics of isolated proteins is important for the functional role of FlgM and some other proteins at high temperatures.

We thank Jian Zhan for helpful discussions. We also acknowledge the support of the Griffith University eResearch Services Team and the use of the High Performance Computing Cluster Gowonda to complete this research.

This work was undertaken with the aid of the research cloud resources provided by the Queensland Cyber Infrastructure Foundation.

We are also indebted to Lukas Folkman for carefully proofreading the manuscript.

This research was supported in part by the National Institute of General Medical Sciences, National Institutes of Health (award number R01GM085003), and grants 61271378, 31000324 and 30970561 from the National Natural Science Foundation of China.

REFERENCES

1. Dunker, A. K., J. D. Lawson, …, Z. Obradovic. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59.
2. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
3. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.
4. Radivojac, P., L. M. Iakoucheva, …, A. K. Dunker. 2007. Intrinsic disorder and functional proteomics. *Biophys. J.* 92:1439–1456.
5. Reichmann, D., Y. Xu, …, U. Jakob. 2012. Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell.* 148:947–957.
6. Babu, M. M., R. W. Kriwacki, and R. V. Pappu. 2012. Structural biology. Versatility from protein disorder. *Science.* 337:1460–1461.
7. Tompa, P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37:509–516.
8. He, B., K. Wang, …, A. K. Dunker. 2009. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19:929–949.
9. Zhang, T., E. Faraggi, …, Y. Zhou. 2012. SPINE-D: accurate prediction of short and long disordered regions by a single neural-network based method. *J. Biomol. Struct. Dyn.* 29:799–813.
10. Fisher, C. K., and C. M. Stultz. 2011. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 21:426–431.
11. Ullman, O., C. K. Fisher, and C. M. Stultz. 2011. Explaining the structural plasticity of α-synuclein. *J. Am. Chem. Soc.* 133:19536–19546.
12. Sethi, A., J. Tian, …, S. Gnanakaran. 2012. Identification of minimally interacting modules in an intrinsically disordered protein. *Biophys. J.* 103:748–757.
13. D'Amico, S., J. C. Marx, …, G. Feller. 2003. Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* 278:7891–7896.
14. Feller, G. 2007. Life at low temperatures: is disorder the driving force? *Extremophiles.* 11:211–216.
15. Feller, G. 2010. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys. Condens. Matter.* 22:323101.
16. Burra, P. V., L. Kalmar, and P. Tompa. 2010. Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS ONE.* 5:e12069.
17. Sabath, N., E. Ferrada, …, A. Wagner. 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* 5:966–977.
18. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein science: a publication of the Protein Society.* 11:739–756.
19. Rauscher, S., and R. Pomes. 2010. Molecular simulations of protein disorder. *Biochem. Cell Biol.* 88:269–290.
20. Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27:527–533.
21. Zhou, Y., and M. Karplus. 1997. Folding thermodynamics of a model three-helix-bundle protein. *Proc. Natl. Acad. Sci. USA.* 94:14429–14432.
22. Zhou, Y., and M. Karplus. 1999. Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis. *J. Mol. Biol.* 293:917–951.
23. Monastyrskyy, B., K. Fidelis, …, A. Kryshtafovych. 2011. Evaluation of disorder predictions in CASP9. *Proteins.* 79 (*Suppl 10*):107–118.
24. Zhang, T., E. Faraggi, …, Y. Zhou. 2013. Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochem. Biophys.* May 31 [Epub ahead of print].
25. Ohnishi, K., K. Kutsukake, …, T. Lino. 1992. A novel transcriptional regulation mechanism in the flagellar regulon of Salmonella typhimurium: an antisigma factor inhibits the activity of the flagellum-specific sigma factor, sigma F. *Mol. Microbiol.* 6:3149–3157.
26. Chilcott, G. S., and K. T. Hughes. 2000. Coupling of flagellar gene expression to flagellar assembly in Salmonella enterica serovar typhimurium and Escherichia coli. *Microbiol. Mol. Biol. Rev.* 64:694–708.
27. Daughdrill, G. W., L. J. Hanely, and F. W. Dahlquist. 1998. The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry.* 37:1076–1082.
28. Dedmon, M. M., C. N. Patel, …, G. J. Pielak. 2002. FlgM gains structure in living cells. *Proc. Natl. Acad. Sci. USA.* 99:12681–12684.
29. Molloy, R. G., W. K. Ma, …, M. J. Gage. 2010. Aquifex aeolicus FlgM protein exhibits a temperature-dependent disordered nature. *Biochim. Biophys. Acta.* 1804:1457–1466.
30. Sorenson, M. K., S. S. Ray, and S. A. Darst. 2004. Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. *Mol. Cell.* 14:127–138.
31. Ma, W. K., R. Hendrix, …, M. J. Gage. 2013. FlgM proteins from different bacteria exhibit different structural characteristics. *Biochim. Biophys. Acta.* 1834:808–816.
32. UniProt Consortium. 2011. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* 39 (Database issue):D214–D219.
33. Huang, S. L., L. C. Wu, …, M. T. Ko. 2004. PGTdb: a database providing growth temperatures of prokaryotes. *Bioinformatics.* 20:276–278.
34. Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302:205–217.
35. Faraggi, E., Y. Yang, …, Y. Zhou. 2009. Predicting continuous local structure and the effect of its substitution for secondary structure in fragment-free protein structure prediction. *Structure.* 17:1515–1527.
36. Faraggi, E., T. Zhang, …, Y. Q. Zhou. 2012. SPINE X: improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles. *J. Comput. Chem.* 33:259–267.
37. Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57:758–771.
38. Altschul, S. F., W. Gish, …, D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
39. Altschul, S. F., T. L. Madden, …, D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
40. Reference deleted in proof.
41. Reference deleted in proof.
42. Dunker, A. K., and Z. Obradovic. 2001. The protein trinity—linking function and disorder. *Nat. Biotechnol.* 19:805–806.
43. Daughdrill, G. W., M. S. Chadsey, …, F. W. Dahlquist. 1997. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat. Struct. Biol.* 4:285–291.
44. D'Amico, S., C. Gerday, and G. Feller. 2001. Structural determinants of cold adaptation and stability in a large protein. *J. Biol. Chem.* 276:25791–25796.
45. Xue, B., R. W. Williams, …, V. N. Uversky. 2010. Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Syst. Biol.* 4 (*Suppl 1*):S1.